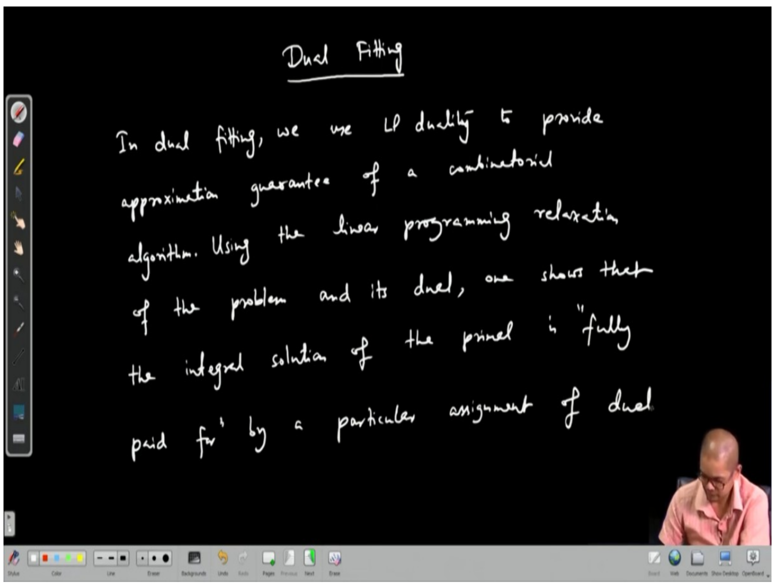
Selected Topics in Algorithm
Prof. Palash Dey
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture - 49
Dual Fitting

Welcome, so in today's class we will see set cover problem again but using linear programming and we will analyse we will provide another analysis of our greedy algorithm for set cover using linear programming this is this approach is called dual fitting.

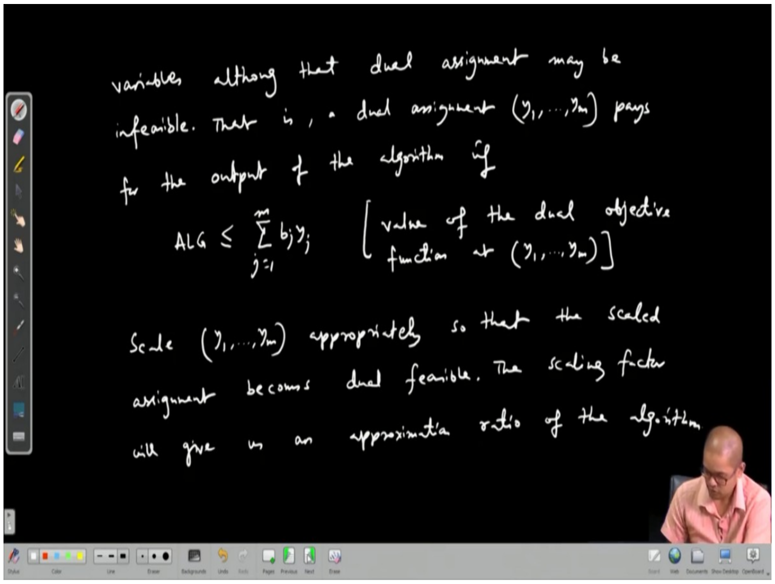
(Refer Slide Time: 00:43)



So, today's topic is dual fitting, so what is the idea of dual fitting? So, this in dual fitting we use LP duality to provide approximation guarantee of a combinatorial algorithm. So, we use a linear program not to directly solve it we use it to analyse it we are not solving linear programs. So, this is falls under second approach of using linear program for approximation algorithm design. So, the idea is using the linear programming relaxation of the problem and its dual.

One shows that the primal integral solution of the integral solution of the primal is fully paid for by a particular assignment of dual variables.

(Refer Slide Time: 04:21)

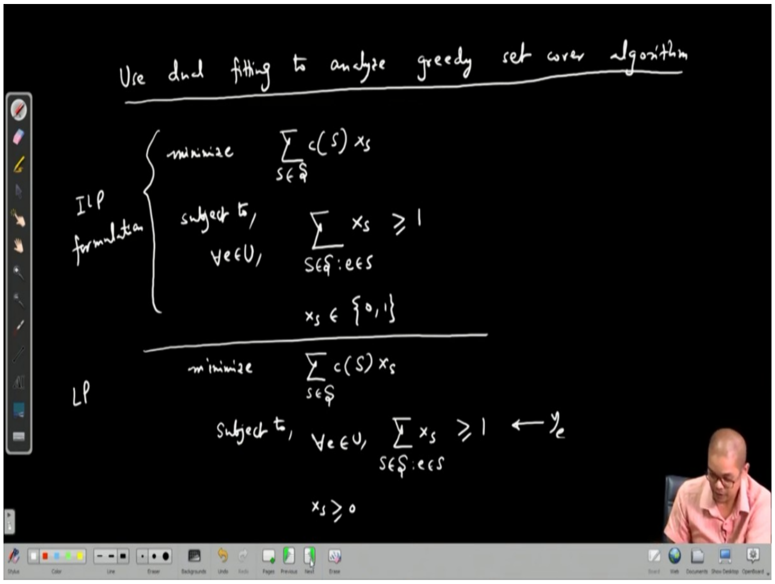


Although that dual assignment may be infeasible, now what do I mean by paid for that means that is a dual assignment say $y_1,\dots,y_m$ pays for the output of the algorithm. So, there is just an algorithm which we are analysing so for the output of the algorithm. If you know ALG, the quality of the solution the value of the optimization function of the primal at the output of the algorithm ALG is a minimization problem.

So, this is ALG is less than equal to $\sum_{j=1}^{m} b_j y_j$. The value of the dual as a dual object function at $y_1,\dots,y_m$, this is value of the dual objective function at $y_1,\dots,y_m$. So, this is the first step and then the second step, the so the first step is to come up with a dual assignment which may be invisible but this gives a bound on ALG and then the second one is scale $y_1,\dots,y_m$ appropriately.

So, that the scaled assignment becomes dual feasible. This the scaling factor will give us an approximation ratio of the algorithm.

(Refer Slide Time: 09:12)



So, what we will do next is we will use dual fitting to analyse the greedy set cover algorithm. So, use dual fitting to analyse greedy set cover algorithm. Why we do that? We already have analysed it using some other approach and we have seen that the approximation ratio is log n but this analysis framework will allow us we will see that to you know to analyse some similar algorithm for some generalization of the set covered problem.

And the analysis will be the analysis of those generalized version of the set cover will be exactly similar that is the gain. So, for this to apply dual fitting what we do is that the first step is to write down the linear integer linear programming formulation of the problem. So, here we will solve a more general set cover problem where each set is illustrated with the cost and the job is to not to pick the minimum number of sets.

It is to pick the set of a collection of sets which minimizes the total cost. So, we need to minimize the total cost of the sets picked. Minimize $\sum c_S x_S$, so we have variable for each set of the set cover instance and c_S is the cost and it is a 0 1 variable. So, if the set S is picked then x_S is 1 otherwise it is 0. So, you are summing over all sets in the collection of sets subject to what are the constraint that each element should be picked.

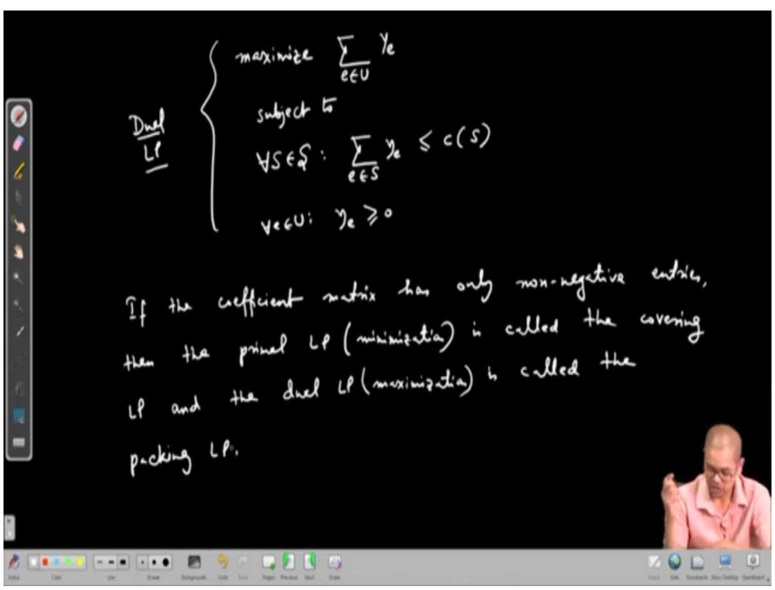
That means if I look at an element these for all element e is in s such that U belongs to e belongs to S among all the sets where e belong at least one of them should be picked. So, x is greater than equal to 1, this is for all is and what is excess x is takes value in 0 1. So, this is the ILP formulation. What we do next is look at its linear programming relaxation. So, LP relaxation is again this objective function remains same minimize $\sum_{s\in S} c_s x_s$ subject to you know.

For each element e in the universe summation x_s in cal S e belong to S is greater than equal to 1 and x_s is greater than equal to 0 and less than equal to 1. Now see that you know because the objective function is monotonically increasing in x is each x is and we want to minimize it we can safely drop the upper bound on x_s and you can write x is equal to greater than equal to 0.

And this way the objective function value does not change by dropping the constraint that x is less than equal to 1, so those constraints are redundant constraints. So, this is the primal linear program LP relaxation of the set cover problem. Now let us write the dual, so for dual what we do is that I have a constraint for each element in the universe. So, I will multiply it with a variable let us call it say y_e.

And the coefficient of each x is you know this which is c s this should give a bound on the coefficient of y_i's x coefficient of x is here after summing all these inequalities. So, what is the goal?

(Refer Slide Time: 14:55)

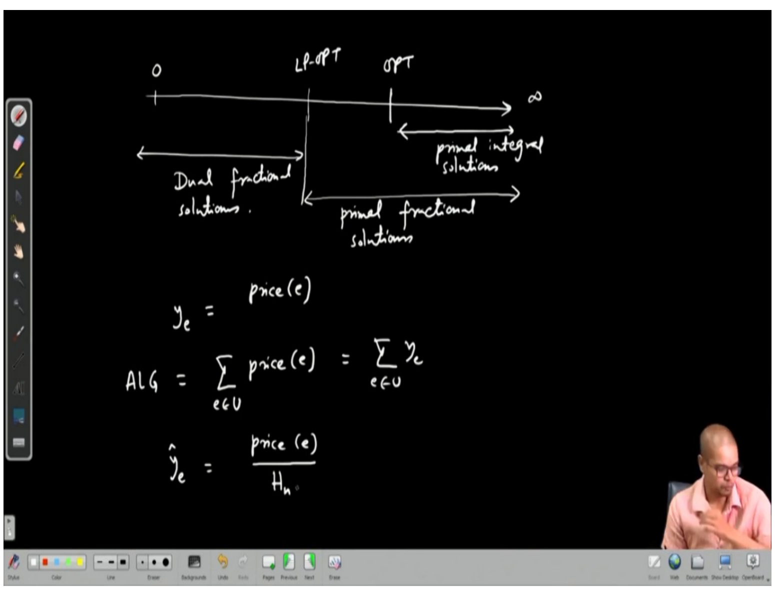


The goal is to maximize, maximize what? I multiply these inequalities and sum all these cardinality inequalities in the right hand side we want to maximize that is $\sum_{e\in U} y_e$ subject to, now look at this set x is which sets which constraints it belongs to all the constraints corresponding to its element. So, the constraint is for all set $s\in S$ $\sum_{e\in S} y_e$ this is less than equal to c_S and we have y_e is greater than equal to 0 for all e in U.

You know this is the dual LP and this one is the primal LP. Now we observe that the primal is a minimization problem and its coefficient matrix all entries are 0 such a program is called the covering program covering LP. On the other hand, the dual is like you know again all constraints are the coefficient matrix has only non-zero entries and it is like a packing problem, so this is called a packing LP.

So, if the coefficient matrix has only non-negative entries, then the primal LP which is a minimization LP is called the covering LP and the dual LP which is a maximization is called the packing LP. Now this LP is often arise in natural problems and that is why they are given some natural in the more names.

(Refer Slide Time: 18:51)

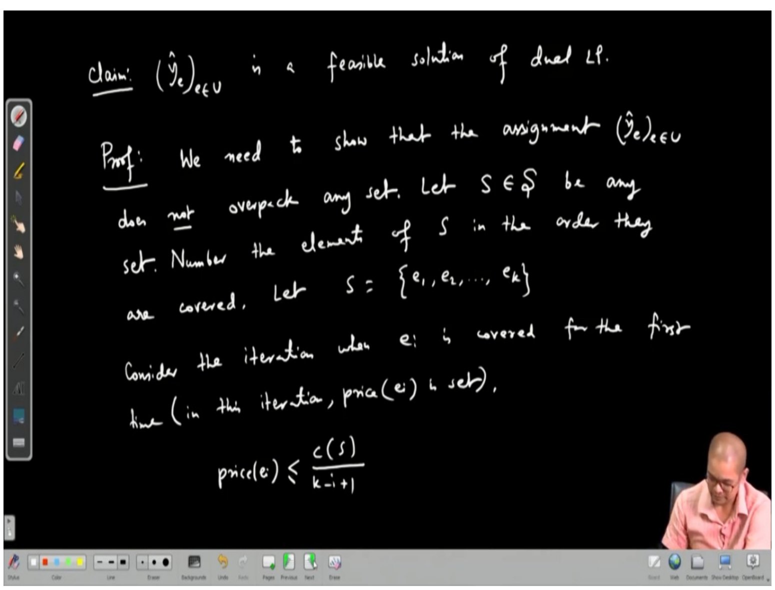


Now if I write down, so if suppose this is here somewhere 0 and infinity so here is suppose the OPT, optimal integral solution this is here this region is primal integral solutions. Now if I relax it this is may be LP OPT, so this region is primal fractional solutions and this region is dual fractional solutions. Now for dual fitting what is the first approach? The first approach is to bound a ALG, the performance of the algorithm with respect to the sum setting of the dual variable.

So, we set you know y_e = price of e, so simply set y = price of e, then ALG we know this is sum of $e \in U$ this is because this we described in the analysis of set cover greedy algorithm for set cover. So this is price of e this is because whenever I pick a set its price has its cost has been distributed equally among the uncovered elements among the elements which are newly covered by this set.

Suppose this set has 10 elements but it when it is picked it some 5 elements is already covered. So, it covers new 5 elements and their prices are set to be the cost of the set by 5, so it is just so the sum of prices of all the elements is the sum of the cost of the sets picked by the algorithm. Hence ALG is sum of prices of this, sum of prices of the elements but this is sum of y_e in U but this will be it is easy to check that this will be an infeasible solution. So, what we define is that we define y e hat is price of e by H_n, the n-th harmonic number.

(Refer Slide Time: 22:40)



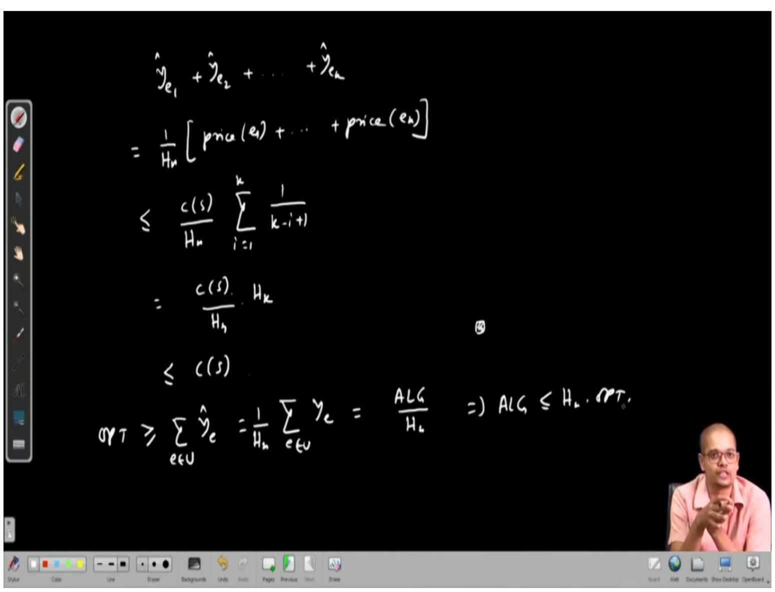
We claim $\hat{y}_e$ e in U is a feasible solution of dual LP. Now what we will achieve by this if it is a feasible solution, it is so it says stays somewhere here so that means this is a lower bound on OPT and using this, that we will show the approximation factor of the algorithm. So, proof, so we need to show that this set of variables satisfies all dual variables all dual constraints that means no set is overpacked.

So, we need to show that the assignment $\hat{y}_e \; e \in U$ does not over pack any set. So, let us pick any set let S belong to the collection be any set, again number the elements in the order they are covered, number the elements of S in the order they are covered. Let S be $e_1, e_2, \ldots, e_k$, these are the k suppose it has k elements $e_1, e_2, \ldots, e_k$. Now consider the iteration when e_i is covered, so we will bound the price of e_i because price of e_i is set when e_i is covered for the first time.

So, consider the iteration when e_i is covered for the first time in this iteration only price of e_i is set. So, whichever set picks it, you know what is the if S picks it if S is used to pick it then you know still in that iteration $e_i, \ldots, e_k$ they are still uncovered. So, if S is picked then S can cover e_i at a cost of C of S by and it covers the remaining $e_i, \ldots, e_k$ elements, so at least $k - i + 1$ elements it will cover at least this many element a new elements it will cover.

So, price of i price of e_i must be less than equal to this, because this element this set S can cover e i at this price. So, if some other set is picked then that can be picked only if it covers e_i at lesser price. So, price of e_i is less than equal to this. So, now what is the constraint?

(Refer Slide Time: 27:48)



So, the constraint here is summation y_e that means summation $y_{e_1}, \ldots, y_{e_k}$ should be less than equal to c_S so that we need to show so $y_{e_1} + \ldots + y_{e_k}$ is a hat this is what this is $\frac{1}{H_n}(price(e_1) + \ldots + price(e_k))$. Now price of we have a bound price of e_i is less than equal to c_S, so this is less than equal to cost of s by H_n summation 1 - k - i + 1, i = 1 to k. So, this is $\frac{c_S}{H_n} H_k$.

Now H_n is at greater than equal to H_k because k is greater less than equal to n this is less than equal to c_S. So, this is a dual feasible solution. So, what do we have then you know OPT is greater than equal to any dual feasible solution. So, what is the dual objective? Summation $\hat{y}_e$, this is summation y hat is nothing but y_e, $e \in U$ $\frac{1}{H_n}$ and summation y_e here this is ALG. So, this is ALG, so in together we have ALG is less than equal to H_n times OPT.

Hence it is a H_n factor approximation it has an ALG factor approximation algorithm. So, in the next class we will use this again dual fitting approach to solve a general generalization version of set cover problem, thank you, so let us stop here.